

# Kent Academic Repository

## Full text document (pdf)

### Citation for published version

Lampis, F. (2016) Forecasting the sectoral GVA of a small Spanish region. *Economics and Business Letters*, 5 (2). pp. 38-44. ISSN 2254-4380.

### DOI

<https://doi.org/10.17811/eb1.5.2.2016.38-44>

### Link to record in KAR

<https://kar.kent.ac.uk/75257/>

### Document Version

Author's Accepted Manuscript

#### Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

#### Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

#### Enquiries

For any further enquiries regarding the licence status of this document, please contact:

[researchsupport@kent.ac.uk](mailto:researchsupport@kent.ac.uk)

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

# Forecasting the sectoral GVA of a small Spanish Region

## Abstract

Our main goal in this paper is to evaluate the point forecasting accuracy of several time series econometric models when applied to a small Spanish region. The variable of interest is the sectoral GVA of the Basque Country. The results support the use of causal models, which outperform univariate models, such as ARMA and SETAR, in forecasting accuracy. The use of a causal model, such as a simple Dynamic Regression model, offers a systematic advantage in the case of a small regional economy for which abundant regional statistical information is available.

**Keywords:** Forecast accuracy, time series models, regional models.

## 1 Introduction

The construction of regional econometric models requires a broader amount of statistical information than what usually exists at a regional level. The absence or lack of statistical information on certain relevant variables, such as import and export, led researchers to focus on a single macroeconomic variable such as the gross domestic product (GDP), the gross value added (GVA), or the unemployment rate. Regional models could be centred on one or few regions and called uni-regional models, or they could be based on all regions of a specific national level and called multi-regional models. Among the extensive literature, we can refer to the recent works of Mayor et al. (2007), Kopoin et al. (2013) and Lehmann and Wohlrabe (2015) for the uni-regional models and to the work of Baltagi et al. (2014) for the multi-regional models.

Another topic that has attracted attention in the literature in the last 30 years is the forecasting accuracy of the regional econometric models (see Patuelli and Mayor, 2014). Indeed, it is not yet clear which econometric methods or approaches are best for achieving good forecasts with regional models. In their literature review, Lehmann and Wohlrabe (2014a) noted several gaps to be filled in the study of regional forecasting. In this paper, we try to address some of the research lines proposed by Lehmann and Wohlrabe (2014a): i) we focus on the forecasting accuracy of regional models using quarterly data for the GVA, ii) we include in our empirical analysis nonlinear time series models, and iii) we take into consideration a small region not

among those usually studied. To our knowledge, nonlinear time series models have not yet been applied to regional models with forecasting purposes, and their use is limited to interpreting and explaining the behaviour of regional variables, such as the unemployment rate (e.g., see Jones et al., 1994).

More specifically, we investigate whether a simple causal approach outperforms the forecasting accuracy of linear and nonlinear univariate models such as ARMA and SETAR. Among the many existing nonlinear time-series models, we focused on the SETAR model as it is the most simple to specify, estimate, and interpret; moreover a broad literature about its forecasting accuracy has been produced in last decades. In this paper, the variable of interest is the sectoral gross value added (GVA) of a Spanish region, the Basque Country. Even though this region has well-developed industrial and services sectors, its weight in the national economy is no greater than 6%. Because any regional modelling is likely to suffer from omitted variables bias due to relevant national variables, it is more reasonable to focus on just one macroeconomic variable of interest. For a small region, forecasts should be performed either with univariate models, as Fullerton et al. (2001) and Fullerton and West (1996) suggest, or using a more complex and structured approach, as Lehmann and Wohlrabe (2014b, 2015) propose.

Our results suggest simple causal models outperform univariate models, either linear or nonlinear. The forecasting exercise is performed on 1-step-ahead and 4-steps-ahead point forecasts using a recursive scheme with an expanding window to approximate a genuine forecasting environment. This approach is based on the recommendations of Terasvirta et al. (2005), Ferrara et al. (2012) and Boero and Lampis (2016), who found it very convenient when the forecasting accuracy of nonlinear models is considered. In the next section, we present the statistical information available for the Basque country, and in the third section, we briefly describe the models estimated and compare the forecasting accuracy of our models.

## 2 Statistical information and variables used

In Spain, the decentralization of the administrative structure in the country has created the need to build a system of regional statistics. The publication of the Spanish Regional Accounts (SRA) was the starting point of several regional studies, but its main limitation was that data were available only on an annual basis. Moreover, the SRA does not provide a comprehensive regional macroeconomic picture, with no

information on public consumption, gross capital formation, or external operations on goods and services. Among various Spanish regions, the Basque Country has the more extended and complete system of regional statistics.

All of the variables used in this study refer to the Basque economy and are withdrawn from the EUSTAT (Statistics Office of the Basque Country); the data used are explained in detail in the Appendix. Data for the sectoral GVA are quarterly and seasonally adjusted and range from the first quarter of 1995 to the fourth quarter of 2014. We use chain volume indexes to avoid any monetary effect on the variables. More specifically, we analyse the GVA and sectors composing it: Agriculture, Construction, Industry and Services. In addition, we analyse the GVA of the sub-sectors of the Services sector: Services in Trade, Hotel and Transport; Services in Public Administrations; and Services in Remaining Activities. Generally, for all of these sectors, there is a period of slow growth from 1995 to 2000, a period of high growth from 2001 to 2007, a strong recession during 2008-2009, and another period of very slow growth for the rest of the sample period. The joint use of the ADF and KPSS tests indicated that almost all of the sectoral GVA are integrated of order one  $I(1)$ ; therefore, the analysis will be performed on the quarterly growth rate. Agriculture GVA resulted in  $I(0)$ , and for this variable, only logarithmic transformation is used<sup>1</sup>.

### 3 Empirical analysis and forecast

In what follows, we estimate three types of models: a SETAR, a seasonal ARMA and a Dynamic Regression model. All the models are estimated recursively, that is, the first estimation is performed using the sample 1995.1-2005.4, and a first set of 1 and 4-quarters ahead forecasts are calculated. Then, each time the models are re-estimated by expanding the sample with one observation, and a new set of forecasts are computed. The process is repeated until the last available data. Since we completely ignore the information embodied in the forecasting period in the specification and estimation stages, these forecasts can be considered as genuine forecasts. Additionally, the models are re-specified at the end of each year, such that the first specification is based on data up to fourth quarter of 2005 and the last specification on data up to fourth quarter of 2013. This approach is inspired to the recommendations of Terasvirta et al. (2005) and Boero and Lampis (2016), who suggest that frequent model re-specification increases the forecast

---

<sup>1</sup> The results of the ADF and KPSS tests are not presented here to conserve space but are available upon request.

accuracy of nonlinear models. Moreover, this procedure allows us to take into account the structural breaks occurred between 2008 and 2009 when the Great Recession hit Spanish economy severely. Dummies variables were included in the models' specifications when needed. For all the models the optimal lag length is selected on the basis of the Akaike information criterion<sup>2</sup>. The ARMA models have been estimated using the classical methodology proposed by Box and Jenkins (1976) using the autocorrelation function as the main instrument to specify the models. The seasonal adjustment performed by EUSTAT was unable to completely remove the seasonality in the data. For many of the specifications adopted, we had to include seasonal autoregressive or moving-average terms.

As in the case of ARMA models, the SETAR models requires a complex structure and tend to assume a high order of lags. This is likely a consequence of some remaining seasonality in the variables, which can only be addressed by the SETAR models using a large number of lags. In general, the use of the SETAR model on quarterly and monthly macroeconomic variables leads to less parsimonious models than the use of simple ARMA models (see for example Granger and Teräsvirta, 1993; Luukkonen and Teräsvirta, 1991; or Teräsvirta et al., 2010). The threshold variable used in a SETAR model is a lagged term of the dependent variable, and in our models, this variable is allowed to change every time the model is re-estimated.

The selection of explanatory variables used for the causal models is based on factors related to the same characteristics of each sectoral GVA, and we tried to build a pure regional model considering only variables from the Basque Country's economy. Initially, we tried to specify some cointegrated models, but both the Engle-Granger test and the Johansen cointegration test rejected the hypothesis of cointegration. This was likely due to the presence of structural breaks and a sample that was too short. Therefore, we opted for using a simpler model, such as the Dynamic Regression (DR). A Dynamic Regression model collapses to a standard dynamic model when error term is white noise and no autoregressive or moving average terms are required, otherwise the error term follows an ARMA process. Almost all explanatory variables were adjusted quarterly and seasonally, while for some of them, a seasonal adjustment and a calendar correction were applied. For the Agriculture Sector, only an explanatory variable has been used: the number of workers employed in the Basque agricultural sector. For all the other sectoral GVA up to five explanatory variables are used.

---

<sup>2</sup> More than 80 different specifications have been used depending on the model, sector and sample used, which are not reported here. Detailed information about models estimated is available upon request.

Statistical information available for the Basque Country gave us the possibility to construct quite complex models, employing several activity indexes for Industry and Construction GVAs. In case of the Services sector and its sub-sectors, we employed variables such as employment, sales index in malls, loans issued by banking sector, discharges in ports, average overnight stay in hotels, enrolment trucks, number of foreign tourists, and number of overnight stays of foreign tourists. Some of the explanatory variables used for other sectors, such as indexes for Industry and Construction and employment in the Services sector, are also used for the whole GVA.

4-steps-ahead forecasts for SETAR models are obtained by applying Monte Carlo simulation methods, so that each point forecast is obtained as the average of 1000 replications, (see Franses and van Dijk, 2000). Standard forecasts are calculated for ARMA and DR models and 1-step-ahead forecast for SETAR model. The root mean squared forecast error (RMSFE) is computed on the whole forecasting period, 36 quarters for the 1-step-ahead forecast and 33 quarters for the 4-steps-ahead forecast. We then calculate the ratio between the RMSFE for each pair of models, that is, between the SETAR model and the ARMA model (Ratio1), between the DR model and the SETAR model (Ratio 2) and between the DR model and the ARMA model (Ratio 3). A value greater than one means that the ARMA model outperforms the SETAR model in the case of the first ratio, that the SETAR model outperforms the DR model in the case of the second ratio, and that the ARMA outperforms the DR model in the case of the third ratio. In Table 1, we report these ratios for each sector, and in the last column, we specify which of the three models has the lower RMSFE.

The results are robust and consistent across all sectors and forecast horizons. DR model outperform univariate models both in case of 1-step and 4-steps-ahead forecasts, systematically. There is only one case where the DR model does not have the better forecasting accuracy, 1-step-ahead forecast of the THT Services GVA. For all the other GVAs the DR model has a clear superiority in performance over its univariate counterparts. On average, most of these gains range from a minimum of 5% to a maximum of 50%, and the superiority of the DR is particularly outstanding in case of the whole GVA with a forecasting accuracy better than SETAR model up to three times. The ARMA model performs better than the SETAR model especially in case of the 1-step-ahead forecast, whilst in case of the 4-steps-ahead forecast the SETAR model has a better forecasting accuracy three out of seven times with respect the ARMA model.

Table 1 : Forecast accuracy

1-step-ahead forecast (sample 2006q1-2014q4)

Sector	Ratio 1	Ratio 2	Ratio 3	lower RMSFE
Agriculture	1.08	0.88	0.94	DR
Construction	1.35	0.73	0.99	DR
Industry	2.00	0.37	0.75	DR
Services WS	1.18	0.71	0.84	DR
Services THT	1.30	0.78	1.01	ARMA
Services PA	1.06	0.88	0.94	DR
Services RA	1.39	0.68	0.95	DR
Whole GVA	2.04	0.31	0.64	DR

4-step-ahead forecast (sample 2006q4-2014q4)

Sector	Ratio 1	Ratio 2	Ratio 3	lower RMSFE
Agriculture	0.68	0.93	0.63	DR
Construction	1.09	0.87	0.94	DR
Industry	1.30	0.64	0.83	DR
Services WS	1.32	0.65	0.86	DR
Services THT	0.85	0.98	0.84	DR
Services PA	0.92	0.92	0.85	DR
Services RA	1.20	0.75	0.91	DR
Whole GVA	2.44	0.32	0.79	DR

WS: Whole Sector; THT: Trade, Hotel and Transport; PA: Public Administrations; RA: Remaining Activities. Ratio 1: ratio between the RMSFEs of the SETAR and ARMA models; Ratio 2: ratio between the RMSFEs of the DR and SETAR models; Ratio 3: ratio between the RMSFEs of the DR and ARMA models.

## 4 Conclusions

For small regions, it is always quite difficult to build econometric models and make forecasts with them; often, simple ARMA models are applied in these cases (see Fullerton et al., 2001; Fullerton and West, 1996). In this paper, we tried to address some of the recommendations suggested by Lehmann and Wohlrabe (2014a) to further explore regional forecasting. Specifically, we have analysed the quarterly sectoral GVA of a small Spanish region, the Basque Country. Data were withdrawn from the regional statistics office of Basque Country, EUSTAT. The analysis was performed separately for each sector using the ARMA model, the SETAR model and the Dynamic Regression model.

The Dynamic Regression model was the only causal model we could apply to our data once the use of cointegrated model was discarded. The results support strongly the use of causal models even in their simple form. The DR model outperforms its univariate counterparts, both in cases of 1-step and 4-steps-ahead, systematically. These results are particularly encouraging for further development of regional statistics even for small regions, as they show such statistical information can be used for forecasting purposes conveniently. As Lehmann and Wohlrabe (2014a, 2014b) show the more available the regional statistics are, the more accurate are the forecasts that could be produced at regional level.

## References

- Baltagi, B., Fingleton, B. and Pirotte A. (2014) Estimating and forecasting with a dynamic spatial panel data model. *Oxford Bulletin of Economics and Statistics*, 76(1):112-138, 2014.
- Boero, G. and Lampis, F. (2016), The forecasting performance of SETAR models: an empirical application. *Bulletin of Economic Research*. doi: 10.1111/boer.12068
- Box, G. and Jenkins, G. (1976) Time Series Analysis: Forecasting and Control. Holden-Day, New York, 1976.
- Ferrara, L., Marcellino, M. And Mogliani, M. (2012) Macroeconomic forecasting during the Great Recession: The return of non-linearity? *Working papers Banque de France*, No. 383.
- Franses, P. and van Dijk, D. (2000) Nonlinear time series models in empirical finance, Cambridge University Press, Cambridge.
- Fullerton, T. and West, T. (1996) Assessing the Historical Accuracy of Regional Economic Forecasts, *Journal of Forecasting*, 15(1):19-36.
- Fullerton, T, Laaksonen, M. and West, T. (2001) Regional multi-family housing start forecast accuracy, *International Journal of Forecasting*, 17(2):171-180.
- Granger, C. and Terasvirta, T. (1993) Modelling Nonlinear Economic relationships, Oxford University Press, Oxford.
- Jones, D., Manning, M., and Stevenson, M. (1994) The Unemployment Rate and the Business Cycle in Britain: An Aggregate and Regional Analysis, *Regional Studies*, 28: 591-604.
- Kopoin, A., Moran, K. and Paré, J. (2013) Forecasting Regional GDP with Factor Models: How Useful are National and International Data? *Economics Letters*, 121(2): 267-270.
- Lehmann, R. and Wohlrabe, K. (2015) Forecasting GDP at the Regional Level with Many Predictors. *German Economic Review*, 16(2): 226-254.
- Lehmann, R. and Wohlrabe, K (2014a) Regional economic forecasting: state-of-the-art methodology and future challenges, *Economics and Business Letters*, 3(4):218-231.
- Lehmann, R. and Wohlrabe, K (2014b) Forecasting gross value-added at the regional level: Are sectoral disaggregated predictions superior to direct ones?, *Review of Regional Research*, 34 (1): 61-90.
- Luukkonen, R. and Terasvirta, T. (1991) Testing linearity of economic time series against cyclical asymmetry, *Annales de l'economie et de statistique*, 20/21:125-142.



- Mayor, M., Lopez, A. and Pérez, R. (2007) Forecasting regional employment with shiftshare and arima modelling, *Regional Studies*, 41:543-551.
- Patuelli, R. and Mayor, M. (2014) Introduction to the Special Issue: Advances in Regional Forecasting, *Economics and Business Letters*, 3(4):191-193, 2014.
- Terasvirta, T., van Dijk, D. and Medeiros, M. (2005) Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination, *International Journal of Forecasting*, 21:755-774.
- Terasvirta, T., Tjostheim, D. and Granger, C. (2010) *Modelling Nonlinear Economic Time Series*, Oxford University Press.

## Appendix

Table 2 : Variables employed

Explained Variables	
Variable	Description
LA	$\log(\text{GVA agricultural})$
DLC	$\Delta \log(\text{GVA Construction}) * 100$
DLI	$\Delta \log(\text{GVA Industry}) * 100$
DLS	$\Delta \log(\text{GVA Services (whole sector)}) * 100$
DLSTHT	$\Delta \log(\text{GVA Services in Trade, Hotel and Transport}) * 100$
DLSPA	$\Delta \log(\text{GVA Services in Public Administrations}) * 100$
DLSRA	$\Delta \log(\text{GVA Services in Remaining Activities}) * 100$

Explanatory Variables	
Variable	Description
DLEA	$\Delta \log(\text{Employment in Agriculture}) * 100$
DLEC	$\Delta \log(\text{Employment in Construction}) * 100$
DLES	$\Delta \log(\text{Employment in Services}) * 100$
DLEPS	$\Delta \log(\text{Employment in Public Sector}) * 100$
DLEFA	$\Delta \log(\text{Employment in Family Activities}) * 100$
DLIPCCW	$\Delta \log(\text{Construction Production Index Civil Work}) * 100$
DLIPCE	$\Delta \log(\text{Construction Production Index Edification}) * 100$
DLIPI	$\Delta \log(\text{Industrial Production Index}) * 100$
LUCPC	$\log(\text{Use of Current Production Capacity})$
LTIP	$\log(\text{Tendency Industrial Production})$
DLSIM	$\Delta \log(\text{Sales Index in Malls}) * 100$
DLL	$\Delta \log(\text{Loans issued by banking sector})$
DLDP	$\Delta \log(\text{Discharges in Ports})$
DLAOS	$\Delta \log(\text{Average Overnight Stay in hotels})$
DLET	$\Delta \log(\text{Enrolment Trucks})$
DLNFT	$\Delta \log(\text{Number of Foreign Tourists})$
DLNOSFT	$\Delta \log(\text{Number of Overnight Stay of Foreign Tourists})$

Employment is the whole number of full-time workers in the relative sector. The IPI has been quartered using fixed averages for each quarter starting from the original monthly data. All of the index variables have the same base year: 2005. For the series available only at a monthly level, a conversion to quarterly frequency was performed by taking fixed averages of the corresponding months.